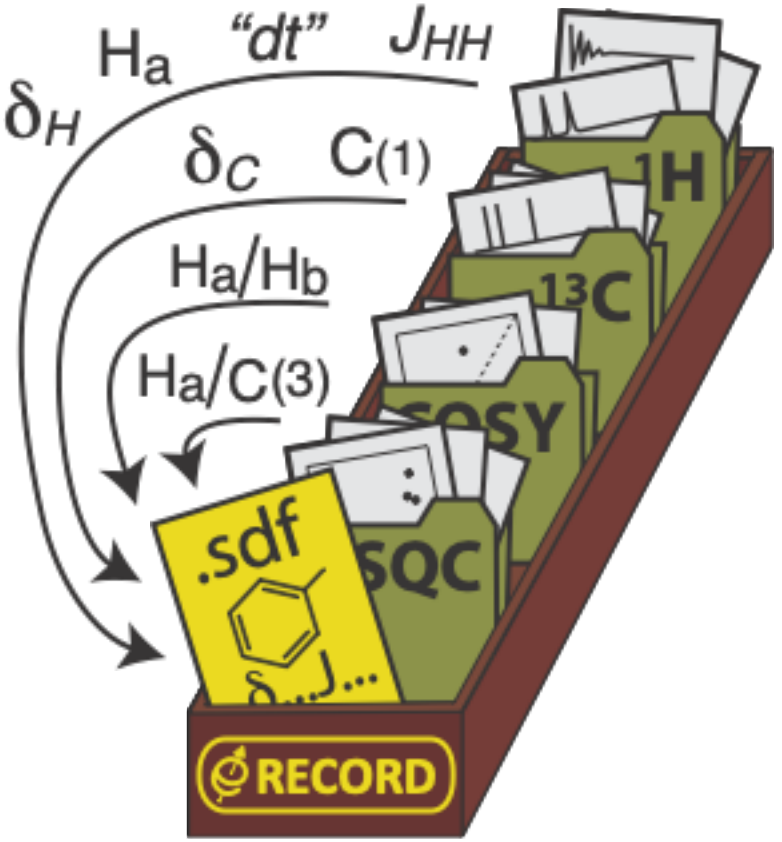
NMReDATA and repo-sitories – making NMR-derived data accessible

The NMReDATA initiative
www.nmredata.org
Stefan Kuhn et al.



Current situation

- Constant progress in NMR methods
- Better resolution, better spectrometers, new experiments, better processing
- Storing and sharing data is lagging behing
- Repositories and databases are one aspect (see ID NMR, NFDI4Chem etc.)
- NMReDATA deals with formats
- Together the FAIR principles can be implemented

FAIR principles

A “set of community-developed guidelines to ensure that data or any digital object are Findable, Accessible, Interoperable and Reproducible. Distinct from peer initiatives that focus on the human scholar, the FAIR principles put a specific emphasis on enhancing the ability of machines to automatically find and use data or any digital object, in addition to supporting its reuse by individuals. Standards for the description, interoperability, citation etc. are at the core of these principles.“ (fairsharing.org)

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. “The FAIR Guiding Principles for scientific data management and stewardship.” Scientific Data 3 (1): 160018. doi:10.1038/sdata.2016.18. http://dx.doi.org/10.1038/sdata.2016.18.

FAIR principles

- FAIRsharing.org is a resource based on the FAIR principles, but it is not the same
- For me, **F**indable and **A**ccesible are mainly about repositories, whereas **I**nteroperable and **R**eusable need formats
- This is a definition of good practices in handling of scientific data. Recently, there is a lot more attention for this area. How far are we in NMR?
- NMReDATA can be the choice for this in NMR

Findable

To be **F**indable:

- F1. (meta)data are assigned a globally unique and persistent identifier

  DOIs are the choice here

- F2. data are described with rich metadata (defined by R1 below)

  Do we have an agreement on metadata in NMR?

Findable

To be **F**indable:

- F3. metadata clearly and explicitly include the identifier of the data it describes

  That‘s easy

- F4. (meta)data are registered or indexed in a searchable resource

  That is one of the tasks of a repository

Accesible

To be **A**ccessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol

  A repository should offer that option

- A1.1 the protocol is open, free, and universally implementable

  Internet infrastructure can be used for this

Accesible

To be **A**ccessible:

- A1.2 the protocol allows for an authentication and authorization procedure, where necessary

  Protection of data before publication/patent is important

- A2. metadata are accessible, even when the data are no longer available

  A repository should do that – institutional repositories do not necessarily do that

Interoperable

To be **I**nteroperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

  This is offered by NMReDATA

- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

  Metadata needs standards

Reusable

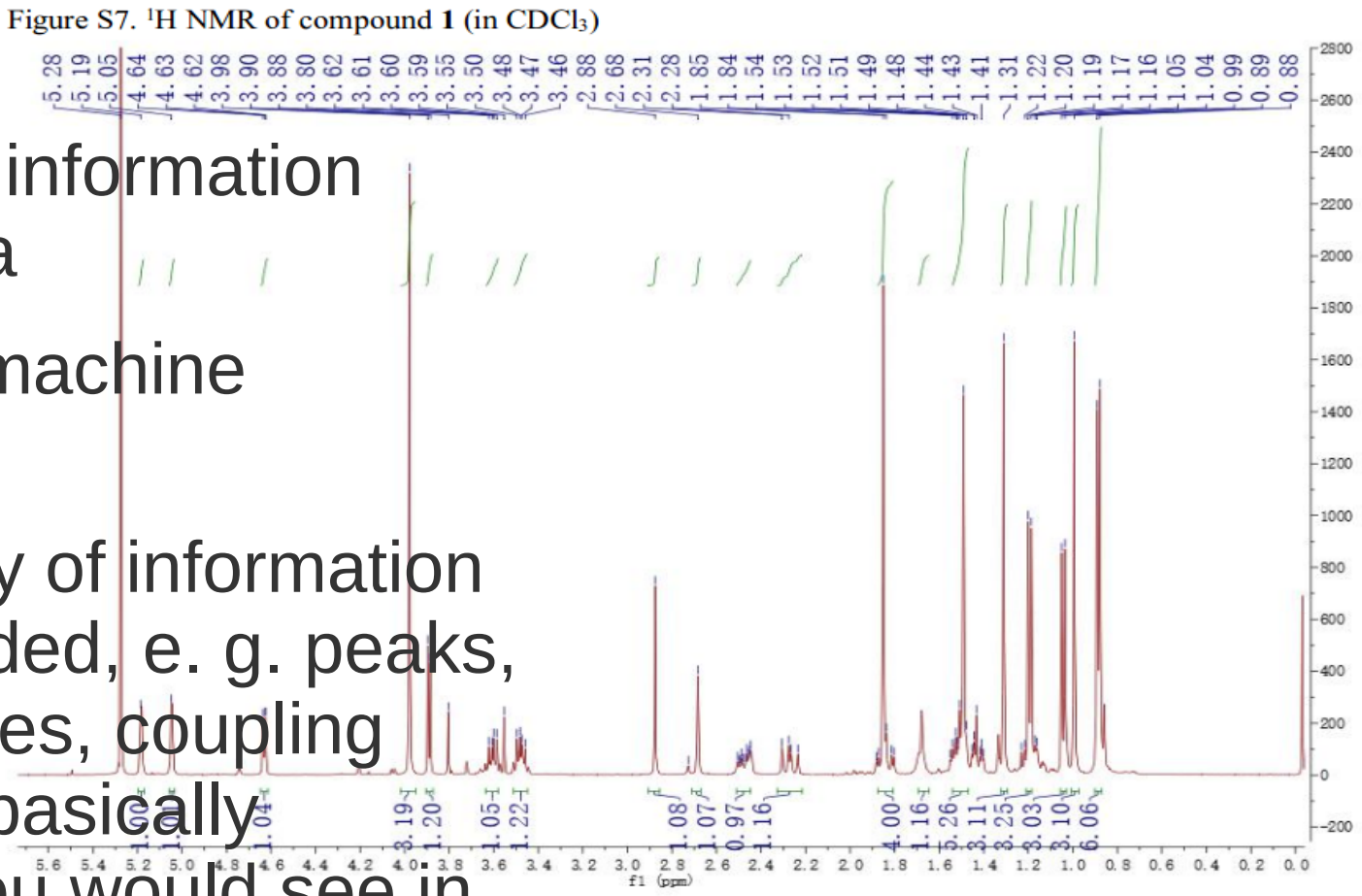
To be **R**eusable:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

This is what a good NMR data format should do

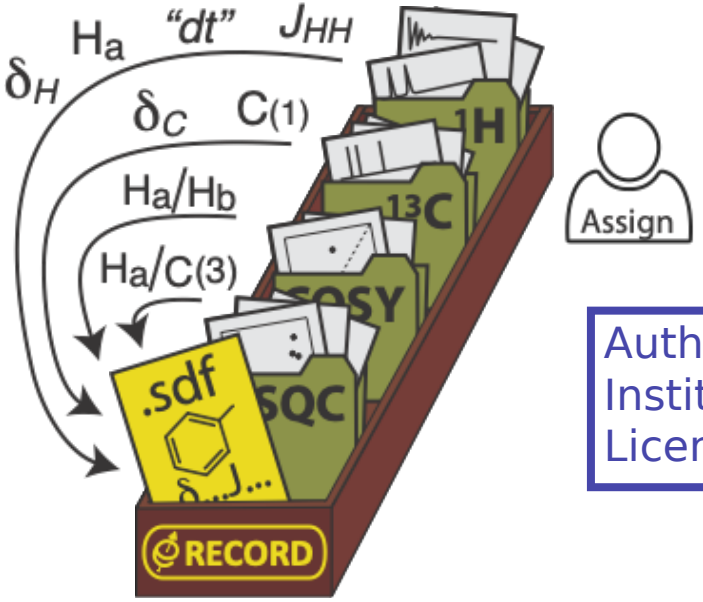
NMReDATA - technical

- NMReDATA is based on the sd file format
- Tags contain information about spectra
- Human and machine readable
- A wide variety of information can be recorded, e. g. peaks, ranges, shapes, coupling constants… basically everything you would see in an annoted spectrum



NMR record - technical

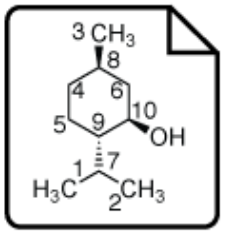
- Combines NMReDATA with raw data in a zip file



Assign

Author
Institution
License

Assigned NMReDATA compound1.nmredata.sdf



NMREDATA_J
H4ax, H4eq, 12.75
H4ax, H5ax, 12.00
H4aq, H5eq, 3.30
...

NMREDATA_ASSIGNMENT
C4, 34.5669, 1
C5, 23.1445, 2
H9ax, 1.1301, 29
C9, 50.1583, 3
H10ax, 3.4302, 30
C10, 71.5891, 4
C6, 45.0568, 5
...

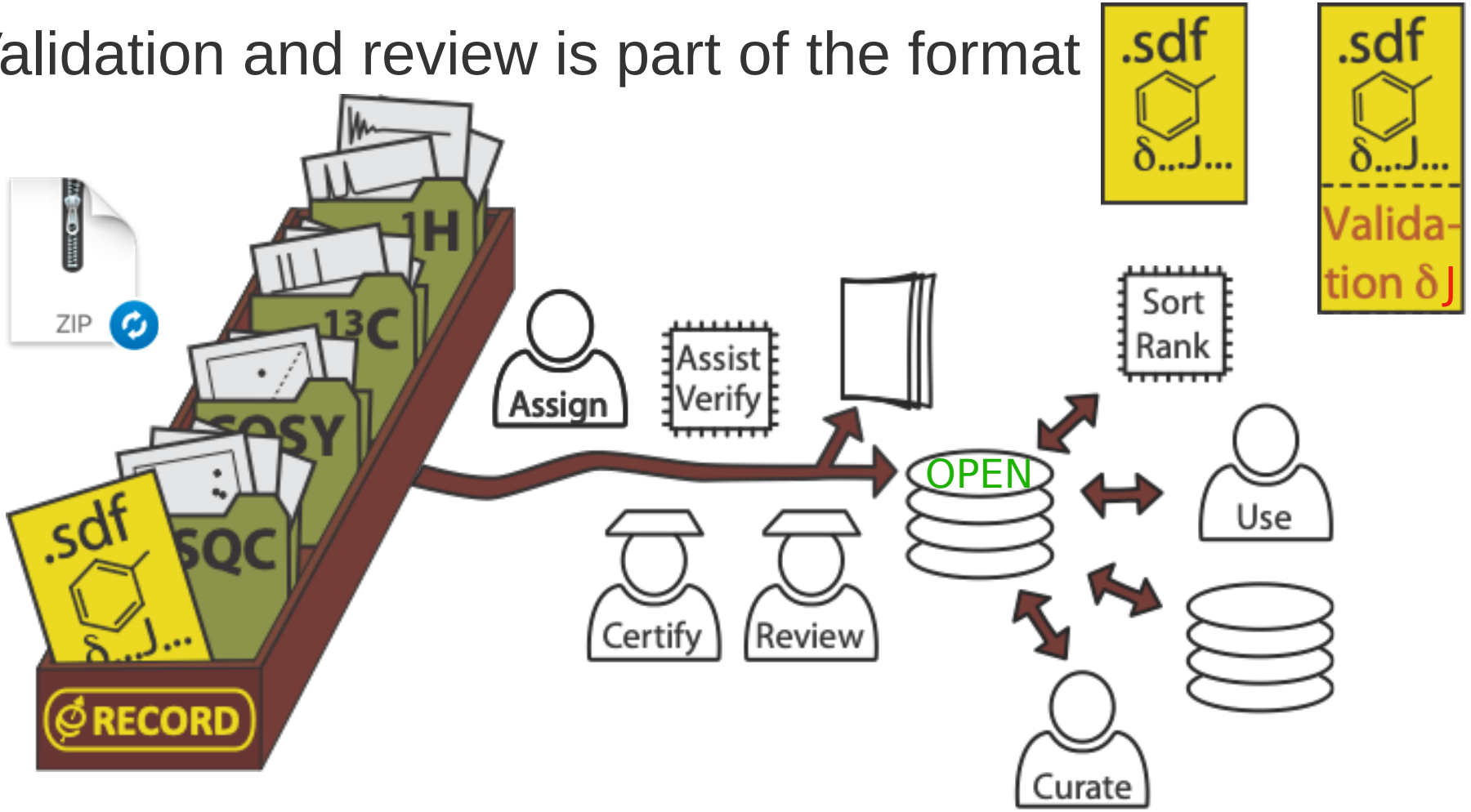
NMREDATA_TEMPERATURE
NMREDATA_SOLVENT
CDCL3

NMREDATA_1D_13C
NMREDATA_1D_1H
1.1301, S=ddd, L=H9ax, E=42.3746, J=12.76, 9.90, 3.05,2.72
3.4302, S=dddd, L=H10ax, E=28.9715, J=10.90,9.90,4.80,4.50
1.4444, S=ddqdd, L=H8ax, E=27.4764, J=12.00,12.00,6.58,4.00,3.00
...

NMREDATA_2D_13C_NJ_1H
NMREDATA_2D_13C_1J_1H
C3/Me3
C4/H4ax
C4/H4eq
...

Validation and review

- Validation and review is part of the format



Certification

- A method to sign (parts of) an NMReDATA file
- Can be used by software to give a quality assurance
- Uses public/private key to ensure authenticity
- Not yet fully specified
- Related to the issue of versioning

Goal

- Have the raw data and all annotiations in one file
- With validation and certification
- Use this for electronic workflows in labs, publication process and for data exchange in software packages

Software support

Working software and tools

Software Platform	Product/tool	Developer	Operating system	Version	NMReDATA	Supported NMReDATA versions	NMR Record	Examples of output
nmrshiftdb2	QuickCheck	Stefan Kuhn	Web browser	4.1.12	generate/write	1.0 & 1.1	generate	link
Javatools GUI and library	see github	Stefan Kuhn	Any Java-enabled	Pre-alpha	generate/write /validate	1.0 & 1.1	generate	
	Spectrus using a script	ACD/Labs	Windows	v. 2017.2 and above	generate/write	1.0	generate/write	
Mnova	Mnova script exporting .mnova files	Damien Jeannerat	Any system running Mnova	V.1.1 (see below)	generate	1.1	generate (When spectra are available)	NMR records

Announced Software & tools

Software Platform	Product/tool	Developer	Operating system	Version	NMReDATA	NMR Record	Examples of output
TopSpin	CMCse	Bruker	Windows, Linux and MAC OS X	V. ?	write	write	
Mnova	Mnova NMR	Mestrelab	Mac OS, Windows, Linux?	Announced for Q1 2018	?		
Matlab/Octave	NMReDATA_View	Damien Jeannerat	Any system running matalab/octave	V0 (will be on GitHub) when ready for test	read	read	
	NOMAD	Tomas Lebl					
Firefox/Google Chrome		Julien Wist	Any running Firefox/Google Chrome		generate	generate	

The end

- Acknowledgements to all collaborators in ID NMR and other projects over the years
- We will have a symposium at SMASH
- The complete revamp of nmrshiftdb2 is on its way
- I am always happy to work on NMR-related software projects

Questions?